

# Exploring the Frontiers of LLMs in Psychological Applications: A Comprehensive Review

Luoma Ke<sup>1</sup>, Song Tong<sup>1,\*</sup>, Peng Cheng<sup>2</sup>, Kaiping Peng<sup>1,\*</sup>

1. Department of Psychology, Tsinghua University

2. School of Social Science, Tsinghua University

\* Corresponding authors: tong.song.53w@kyoto-u.jp; pengkp@tsinghua.edu.cn

## Abstract

This paper explores the frontiers of large language models (LLMs) in psychology applications. Psychology has undergone several theoretical changes, and the current use of Artificial Intelligence (AI) and Machine Learning, particularly LLMs, promises to open up new research directions. We provide a detailed exploration of how LLMs like ChatGPT are transforming psychological research. It discusses the impact of LLMs across various branches of psychology, including cognitive and behavioral, clinical and counseling, educational and developmental, and social and cultural psychology, highlighting their potential to simulate aspects of human cognition and behavior. The paper delves into the capabilities of these models to emulate human-like text generation, offering innovative tools for literature review, hypothesis generation, experimental design, experimental subjects, data analysis, academic writing, and peer review in psychology. While LLMs are essential in advancing research methodologies in psychology, the paper also cautions about their technical and ethical challenges. There are issues like data privacy, the ethical implications of using LLMs in psychological research, and the need for a deeper understanding of these models' limitations. Researchers should responsibly use LLMs in psychological studies, adhering to ethical standards and considering the potential consequences of deploying these technologies in sensitive areas. Overall, the article provides a comprehensive overview of the current state of LLMs in psychology, exploring potential benefits and challenges. It serves as a call to action for researchers to leverage LLMs' advantages responsibly while addressing associated risks.

**Keywords** Large language models (LLMs) • ChatGPT • Machine learning • Artificial intelligence (AI) • Psychology • Research methodology

## 1. Introduction

Artificial intelligence (AI) has a history spanning nearly seven decades, beginning with the 1956 Dartmouth Conference. The field has recently been revolutionized by the advent of large language models (LLMs), such as OpenAI's ChatGPT series, Google's Bard, and Meta's Llama. Notably, OpenAI's GPT-4, in particular, may signify a paradigm shift by demonstrating impressive capabilities (e.g., it solved difficult

tasks in math, coding, vision, medicine, law, psychology, etc.) (Bubeck et al., 2023), that is to say, AI for Science (H. Wang et al., 2023). LLMs mark a critical juncture in the evolution of machine learning and AI, propelled by their expansive size and sophisticated neural architectures that incorporate attentional mechanisms (Vaswani et al., 2017). Due to the integration of cognitive mechanisms (Binz & Schulz, 2023a), these models have acquired the ability to exhibit emergent behaviors akin to complex physical systems (Wei et al., 2022), which has not only enhanced their ability to understand concepts and high-level semantics (J. Li et al., 2022) but also deepened our insights into cognitive processes (Sejnowski, 2022). In psychological applications, these developments are reshaping interactions and comprehension of data, language, and our environment (De Bot et al., 2007; Demszky et al., 2023), contributing significantly to various fields, including clinical (Thirunavukarasu et al., 2023), development (Frank, 2023; Hagendorff, 2023), and social psychology (Demszky et al., 2023; Hardy et al., 2023; Zhang et al., 2023). Moreover, they profoundly impact psychological research methodologies, offering novel approaches and tools for exploration and analysis.

## **1.1. The concept of LLMs: From machine learning to capability emergence**

Machine learning, particularly natural language processing (NLP), has significantly progressed in the last decade. However, the emergence of LLMs such as GPT-4 and its predecessors marks a significant leap in AI capabilities. LLMs are deep learning models designed to process natural language text and generate human-like responses or texts. Their capability emergence is defined as a qualitative change in behavior due to a quantitative change in the system and thus a qualitative change in behavior, i.e., a capability is emergent if it does not exist in a smaller model and exists in a larger model (Wei et al., 2022).

At the heart of this LLM is the transformer architecture, a deep neural network with an attentional mechanism to efficiently process sequential data in parallel (Vaswani et al., 2017), which works somewhat similarly to the human brain functions. This architecture has revolutionized the field of Natural Language Processing (NLP). The self-attention mechanism of the transformer architecture captures contextual relationships in textual data, allowing for more sophisticated language understanding. On the one hand, the "large" in LLMs refers to many parameters and massive amounts of training data used to fine-tune these models, typically billions of parameters and terabytes of text (Binz & Schulz, 2023b), and on the other, means master the world model (Yildirim & Paul, 2023).

The process of large language modeling, from machine learning to the emergence of competence, can be divided into several key stages (Demszky et al., 2023). First, pre-training: LLMs are pre-trained on large

amounts of textual data to learn intricate linguistic, syntactic, and textual structures (P. Liu et al., 2023). This unsupervised learning phase lays the foundation for the big language model to understand the language. Second, fine-tuning: LLM can be fine-tuned for a specific task or domain after pre-training to make it adaptable and suitable for various applications (Liu et al., 2022). This fine-tuning process ensures that the model can generate contextually relevant responses and engage in meaningful conversations or tasks. Third, language comprehension: LLMs have demonstrated a remarkable ability to understand and develop human-like text. They can answer questions, write articles, summarize content, translate language, and even do creative writing (Bubeck et al., 2023). Their skillful understanding of context is an essential factor contributing to the emergence of their intelligence. Fourth is the emergence of capabilities: LLMs exhibit "capability emergence" when integrated into various applications and systems. They can perform tasks that require a deep understanding of language and context, often achieving human-like or superhuman performance in specific domains (OpenAI, 2023), such as analogical reasoning (Webb et al., 2023), creativity (Stevenson et al., 2022), and emotion recognition (Patel & Fan, 2023).

Therefore, LLMs offer intriguing insights into how these technologies can mimic or augment human cognitive processes. For instance, LLMs' ability to understand and generate natural language echoes aspects of human linguistic and cognitive skills (Goertzel, 2023). This parallel allows for exploration into AI applications in the cognitive psychology (Sartori & Orrù, 2023), language acquisition (Jungherr, 2023), and even the mental health (Lamichhane, 2023). Moreover, the study of LLMs contributes to our understanding of the human mind, offering a computational perspective on language processing, the decision-making (Sha et al., 2023), and learning mechanisms (Hendel et al., 2023). Fusing these disciplines advances AI's capabilities and deepens our comprehension of the human mind.

## 1.2. Psychology and AI

Psychology, as a science that explores the human mind and behavior, has undergone significant theoretical changes since the late 19th century, with psychoanalysis and behaviorism extending to cognitive psychology (Hothersall & Lovett, 2022). This history not only marks a shift in the focus of research in psychology but also reflects the academic trend from observing behavioral manifestations to exploring in-depth psychological connotations. Each of these phases has led to a deepening understanding of the psycho-cognitive processes of human beings.

Understanding human psycho-cognitive processes is therefore crucial to psychology. In clinical and

counseling psychology, research in cognitive psychology supports diagnosing and treating psychological disorders. It deepens understanding of the psychological mechanisms underlying emotions, stress, and human behavior. Psychotherapies, such as cognitive-behavioral therapy (Hofmann et al., 2012) and psychodynamic therapy, have become essential tools for promoting mental health and emotional regulation. In educational and developmental psychology, the development of cognitive psychology has fostered a deeper understanding of the role of perceptual and affective factors in the learning process (Glaser, 1984), which has led to innovations in teaching methods and learning strategies. In social and cultural psychology, cognitive psychology research helps explain individuals' behavior and mental processes in different social and cultural contexts. It explores how cultural differences affect individuals' cognitive patterns, values, and behavioral norms, especially in globalization, interaction, and integration. Meanwhile, in social psychology, cognitive psychology's research on group behavior, social influence, prejudice, and discrimination is of great value in promoting social harmony and mutual understanding (Park & Judd, 2005).

AI is a growing force in psycho-cognitive research. For example, Simon (1979) recognized the potential of computational models to simulate human cognitive processes early on. The recent emergence of LLMs, represented by OpenAI's GPT family (mainly including GPT-3, ChatGPT, and GPT-4), which can process and generate human-like texts and perform close to humans in some cognitive tasks (Bubeck et al., 2023). More so, it provides a unique perspective to study human cognition. For example, GPT-3 can solve vignette-based tasks similarly or better than human subjects and can make rational decisions based on descriptions, outperforming humans in the multi-armed bandit task (Binz & Schulz, 2023b). Furthermore, after extensive testing, GPT-3 can solve complex analogical problems at levels comparable to human performance, and analogical reasoning is an essential hallmark of human intelligence (Webb et al., 2023). Furthermore, fine-tuning across multiple tasks could allow LLMs to predict human behavior in previously unseen tasks, i.e., LLMs could be adapted to general-purpose cognitive models (Binz & Schulz, 2023a), potentially opening up new research directions that could transform cognitive psychology and behavioral science in general.

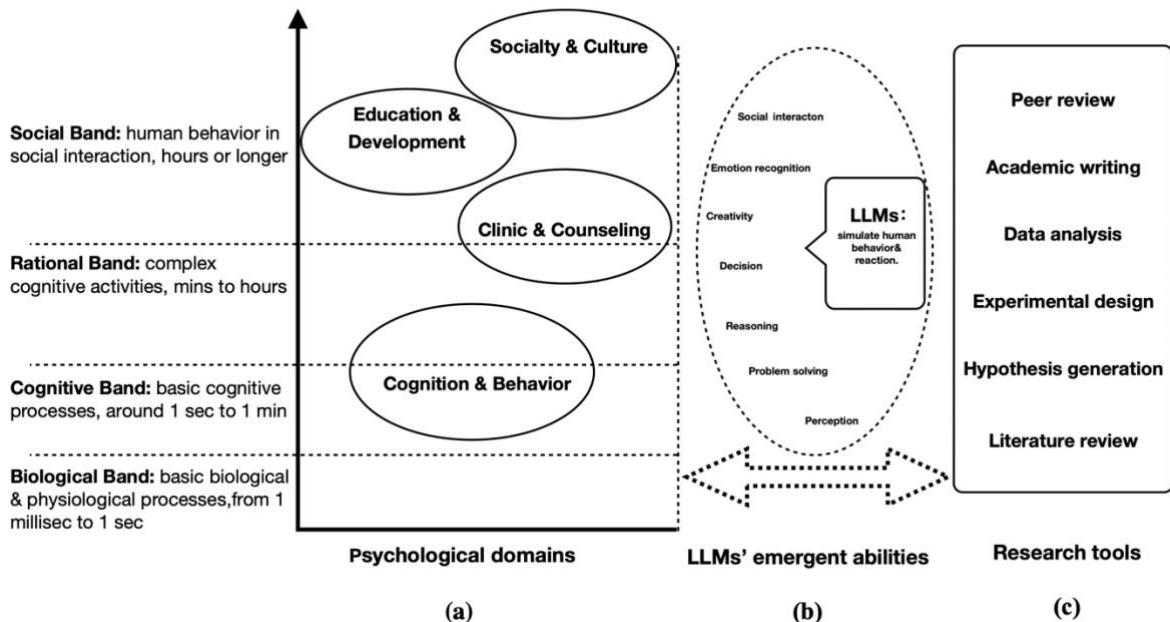
Newell's time-scale theory provides a multidimensional framework for understanding human behavior Newell (1990). In his seminal work, Newell (1990) articulates a nuanced framework for comprehending cognitive processes, stratifying human behavior across four distinct temporal levels (see Fig. 1a). At the foundational biological level, he addresses core biological and physiological processes characterized by rapid time scales fluctuating from approximately one millisecond to one second. This level might include neural responses and sensory processing, foundational to human cognition. Advancing to the cognitive layer,

Newell examines fundamental cognitive mechanisms operating on intermediate time scales, typically from one second to around one minute. This layer could encompass basic cognitive operations such as attention, perception, and short-term memory. Further, at the rational layer, the focus shifts to more elaborate and sustained cognitive activities. These processes, often extending from several minutes to a few hours, might involve complex problem-solving, planning, and decision-making activities requiring a higher level of cognitive engagement. Finally, the social layer encapsulates human behavior within social interactions and structures. This level, characterized by the broadest time scales ranging from a few hours to days or more, delves into the dynamics of social communication, group behavior, and cultural influences on cognition. This layered approach underscores the multifaceted nature of human behavior, highlighting the interplay between rapid physiological processes and the more prolonged, socially influenced aspects of human cognition.

LLMs have great potential for modeling cognition and behavior on these different time scales (see Fig. 1b) and can provide new insights into human psycho-cognitive processes. Recent research has revealed significant advancements in LLMs' ability to emulate complex human cognitive and social behaviors (Grossmann et al., 2023; Marjeh et al., 2023; Orru et al., 2023; Pal et al., 2023; Stevenson et al., 2022; Webb et al., 2023). Grossmann et al. (2023) and Marjeh et al. (2023) have shown LLMs' proficiency in simulating human social interactions and perceptual processing, respectively. Orru et al. (2023) and (Webb et al., 2023) highlighted LLMs' capabilities in complex problem-solving and reasoning, while Hagendorff et al. (2023) focused on decision-making processes. Stevenson et al. (2022) documented LLMs' potential for creativity, and Patel and Fan (2023) demonstrated their ability in emotion recognition. These findings collectively indicate the expanding role of LLMs in mimicking and enhancing human cognitive and social functions, marking significant progress in AI research.

As a general cognitive model (Binz & Schulz, 2023a), LLMs can provide new perspectives and approaches to research in the fields of cognitive and behavioral psychology, clinical and counseling psychology, educational and developmental psychology, and social and cultural psychology in different time scales of human behavior (see Fig. 1a).

Newell's (1990) Time scales of human behavior



**Fig.1 LLMs' emergent abilities can be applied in psychological domains and work as research tools: (a) LLMs' application in psychological domains from time scales of human behavior. (b) LLMs' emergent abilities. (c) LLMs work as research tools.**

Based on these emergent abilities, LLMs can also be used as a research aid (see Fig. 1c) to help psychologists with everything from literature review (Aydin & Karaarslan, 2022; Qureshi et al., 2023), experimental subjects (Dillion et al., 2023; Hutson, 2023), and data analysis (Patel & Fan, 2023; Peters & Matz, 2023; Rathje et al., 2023) to academic writing (Dergaa et al., 2023; Stokel-Walker, 2022) and peer review (Chiang & Lee, 2023; Van Dis et al., 2023). Thus, LLMs can potentially become research assistants for psychologists, helping them improve their research efficiency.

### 1.3. Objectives and significance of the present review

This review systematically examines the use of LLMs in various psychological domains, analyzing their applications over different behavioral time scales. The exploration begins with LLMs in cognitive and behavioral psychology (Section 2), followed by their roles in clinical and counseling psychology (Section 3). The review then transitions to educational and developmental psychology (Section 4) and social and cultural psychology (Section 5), outlining LLMs' contributions in each area. To gain a deeper understanding of the impact of LLMs on psychological research, Section 6 will provide an overview of their potential as a tool for scientific research. The review also addresses challenges and future research directions in applying LLMs to psychological contexts. It concludes by summarizing their applications in psychology and offering recommendations for future work. Crucially, this review proposes integration strategies for LLMs in

psychological research and provides insights into interpreting these models from a psychological standpoint, contributing to their safety and interpretability.

## 2. LLMs in cognitive and behavioral psychology

Within multilevel time scales of human behavior (Newell, 1990), cognitive and behavioral psychology has focused primarily on the study of cognitive processes on sub-hourly time scales (see Fig. 1), which encompass humans engaging in perception, memory, thinking, decision-making, problem-solving, and conscious planning. Cognitive and behavioral psychology typically employs experimental methods to study these cognitive processes by controlling and observing behaviors and responses under specific conditions. The recent emergence of LLMs has reinvigorated the discussion as to whether human cognitive abilities are revealed in these LLMs given sufficient training data. If the answer is yes, then it would be possible to study the cognitive processes of LLMs, thereby gaining knowledge of human cognitive processes and forming a valuable addition to existing research methods in cognitive psychology.

Binz and Schulz (2023a) found that fine-tuning multiple tasks enabled the LLM to predict human behavior in previously unseen tasks, suggesting that the LLM can be adapted to become a generalist cognitive model. In another study, they tested the GPT-3 with tools from cognitive psychology and showed that it made better decisions and outperformed humans in a multiarmed bandit task (Binz & Schulz, 2023b). In addition, there are other series of studies that have demonstrated that LLMs have perceptual judgment (Marjieh et al., 2023), reasoning (Webb et al., 2023), and decision-making abilities (Hagendorff et al., 2023), creativity (Stevenson et al., 2022), and problem-solving (Orru et al., 2023), and one study even demonstrated that the GPT-4 has the mental abilities of a seven-year-old child through a false belief task (considered the gold standard for testing theory of mind in humans) (Kosinski, 2023). For example, Hagendorff et al. (2023) explored reasoning capabilities and decision-making processes of the OpenAI GPT family by the following experimental method: Design a series of semantic illusion and cognitive reflection tests designed to elicit intuitive but erroneous responses. Apply these tasks, traditionally used to study human reasoning and decision-making, to OpenAI's family of generative pre-trained Transformer models. Analyze model performance on a Cognitive Reflection Test (CRT) task and a semantic illusions task to reveal their System 1 and System 2 thought processes. Observe how ChatGPT models show correct responses in these tasks and avoid pitfalls. Evaluate the performance of the models in the CRT task by preventing them from chain-

thinking to reason. The results show that as model size and language capability increase, the OpenAI family of generative pre-trained Transformer models increasingly exhibit human-like intuitive system 1 thinking and associated cognitive errors. Table 1 summarizes the applications of LLMs to cognitive and behavioral psychology.

These research cases demonstrate that LLMs have human-like cognitive abilities (Zhuang et al., 2023). Studying the cognitive mechanisms of LLMs would provide new insights into human cognitive processes. They will provide promising avenues for advancing psychological research methodologies and understanding complex cognitive phenomena as they evolve.

**Table 1** Applications of LLMs in cognitive and behavioral psychology study.

Author	Research question	Research method	Key finding
Sartori and Orrù (2023)	The human-like properties LLMs exhibit in a variety of cognitive tasks.	Decision-making, information search, deliberation, causal reasoning, Wason Selection Task, and Raven-like matrices.	LLMs have demonstrated human-like performance in cognitive psychology.
Hagendorff et al. (2023)	Reasoning capabilities and decision-making processes of the OpenAI GPT family bear any resemblance to human system 1 and system 2 thought processes.	Analyze model performance on a Cognitive Reflection Test (CRT) task and a semantic illusions task to reveal their System 1 and 2 thought processes.	ChatGPT-3.5 and 4 use input-output context windows during chain-thinking reasoning, similar to how humans use laptop-support system 2 thinking.
Hutson (2023)	Can AI language models be used to replace human participants in experiments?	LLMs(e.g., GPT-3.5) were used to conduct the experiment instead of human participants.	LLMs can replace human participants in experimental research in some cases.
Dillion et al. (2023)	Explore whether LLMs can replace human participants in the psychological sciences.	Making human-like moral judgments was assessed by analyzing the similarity of GPT-3.5's judgments to humans.	LLMs can be used as a substitute for human participants in some cases.
Zhuang et al. (2023)	How to efficiently measure the cognitive abilities of LLMs.	A Computerized Adaptive Testing (CAT) for assessing cognitive ability in LLMs.	ChatGPT has surpassed the programming abilities of high-ability college students in dynamic programming and search.
Grossmann et al. (2023)	How to improve social science research methods in the context of the ongoing impact of LLMs on social science research.	LLMs replace human participants in data collection, and act as "peers" in social interaction studies.	LLMs have great potential for use in social science research because of their ability to model human behavior and generate diverse responses.
Loconte et al. (2023)	Neuropsychological evaluations of the performance of a LLM in terms of prefrontal functioning.	The Verbal Reasoning Test, Cognitive Estimation task, Metaphor, and Idioms Comprehension test, Winograd Schema, Tower of London, Hayling Sentence Completion Test, Compound Remote Associate problems, and Social Cognition.	ChatGPT exhibits disjointed cognitive profiles in prefrontal functioning (e.g., some performing better than average and others performing at pathological levels).
Binz and Schulz (2023a)	How to better describe human decision-making behavior by fine-tuning LLMs.	By comparing the goodness-of-fit of different models: random guessing model, domain-specific model, LLaMA unfinetuned, and fine-tuned model.	LLMs using fine-tuning (e.g., LLaMA) can successfully capture human decision-making behavior and perform better than domain-specific models.
Orru et al. (2023)	The potential of ChatGPT as an intelligent tool for problem-solving.	Verbal insight problems were administered to ChatGPT: the first set was referred to as "practice problems," while the second set was referred to as "transfer problems".	ChatGPT's global performance in the practice and transfer problems was identical to the most likely results in the human sample.
Hagendorff (2023)	How to use psychological methods to study the emergent abilities and	Studying the behavioral patterns, emergent abilities, and decision-making mechanisms of LLMs by treating them as participants in a	Uncovering emergent abilities in LLMs that cannot be detected by traditional natural language processing

Dhingra et al. (2023)	behaviors of LLMs. The performance of the GPT-4 on a cognitive psychology task to understand how it processes.	psychological experiment. Evaluating the performance of the GPT-4 on a range of cognitive psychology datasets: CommonsenseQA, SuperGLUE, MATH, and HANS.	benchmarks. The GPT-4 has revolutionary potential to help machines bridge the gap between human and machine reasoning.
Shiffrin and Mitchell (2023)	Decision-making mechanisms and other psychological qualities of LLMs.	Two sets of experiments were conducted, using situational prompts from the psychological literature and prompts not in the GPT-3 training corpus.	GPT-3 outperforms humans on some tasks and performs poorly on others.
Binz and Schulz (2023b)	Assessing the GPT-3's cognitive ability.	Vignette-Based Investigations. Decision-Making. Information Search. Deliberation. Causal Reasoning.	GPT-3 showed surprising abilities in decision-making, information search, and thoughtfulness, but performed poorly in causal reasoning.
Marjeh et al. (2023)	How LLMs predict human sensory judgments across multiple sensory modalities.	Using LLMs such as GPT-3, GPT-3.5 and GPT-4 to predict human judgments in six sensory modalities (pitch, loudness, color, sound, taste, and timbre).	LLMs can successfully predict human perceptual judgments in six modalities.
Huang and Chang (2022)	How to improve and direct the reasoning of these models.	Methods for assessing the reasoning ability of LLMs: fully supervised fine-tuning, cueing and contextual learning, problem decomposition, mixed-methods.	Improving the reasoning ability of LLMs requires the use of training data, model architectures, and optimization goals specific to reasoning.
Hagendorff et al. (2022)	The machine intuitive capabilities of human-like intuitive decision-making in GPT-3.5.	Conducting Cognitive Reflection Test (CRT) and Semantic Illusion Test on GPT-3.5.	GPT-3.5 systematically exhibits "machine intuition", i.e., produces human-like erroneous decisions on the CRT as well as semantic illusions.
Stevenson et al. (2022)	Evaluating the performance of Open AI's generative natural language model GPT-3 on creativity.	Comparing the GPT-3 to the Alternative Uses Test (AUT), which is widely used by humans in creativity research.	On the creativity test, humans currently outperformed the GPT-3 in originality and unexpectedness, but the GPT-3 performed better in utility.

### 3. LLMs in clinical and counseling psychology

In multilevel time scales of human behavior (Newell, 1990), clinical and counseling psychology would involve the assessment of everyday behavioral acts (about a few hours to a day), habitual thinking (about a day to a few months), and psychological disorders, among others (a few months to many years) (see Fig.1). Clinical and counseling psychology focuses on assessing, diagnosing, treating, and preventing individual mental health problems. These processes often involve medium- to long-term periods. According to related reports, there is a public rush to use LLMs such as the ChatGPT for mental health screening and treatment (Demszky et al., 2023). LLMs are expected to be used in clinical and counseling because they can parse human language and generate human-like responses, categorize text, and flexibly adapt conversational styles representing different theoretical orientations (Stade et al., 2023). So, how do LLMs work in psychotherapy, and can they replace the role of the human psychotherapist?

LLMs are a basic generalized model with the ability to learn from small samples (Brown et al., 2020), a capability that allows them to quickly become experts in the clinical and counseling domain with only a small amount of data to learn from. For example, LLMs trained on clinical content can identify more specific factors of change that can help psychologists understand the process of clinical interventions, thus opening

the black box of psychotherapy (Schueller & Morris, 2023). Additionally, related studies have shown that LLMs can correctly recognize emotions and respond to (Patel & Fan, 2023; Schaaff et al., 2023) appropriately and that human-computer collaborations in clinical psychological support result in more empathy (Sharma et al., 2023). Also, LLMs can accomplish mental health assessments (Elyoseph & Levkovich, 2023; Kjell et al., 2023) and individualized interventions (Blyler & Seligman, 2023a, 2023b). Blyler and Seligman (2023a) proposed an individualized intervention: Participants were recruited from previous studies and were 18 or older. From the five narrative identities generated by ChatGPT-4 that were rated as "completely accurate," five participants representing different backgrounds and experiences were selected. The participants' narratives were provided through a dialog with the ChatGPT-4, and the AI was asked how it would guide life coaching based on the narrative identities. Based on the AI-generated narrative identities and recommended coaching methods, ChatGPT-4 was asked how to recommend specific interventions. The results suggest that the coaching strategies and interventions generated by ChatGPT-4 make perfect sense based on narrative identity. Table 2 summarizes the applications of LLMs to clinical and counseling psychology.

These research cases, which demonstrate the ability of LLMs to provide clinicians with adequate mental health support (Schueller & Morris, 2023), hold promise to address the lack of capacity in the mental health care system as it continues to evolve and may provide more individualized treatment services and even have the potential to fully automate psychotherapy in the future (Stade et al., 2023). Of course, it is essential in this process to ensure that LLM is safe and privacy-protective in psychotherapy.

**Table 2** Applications of LLMs in clinical and counseling psychology study.

Author	Research question	Research method	Key finding
Carlbring et al. (2023)	Can AI be utilized to improve the effectiveness of Internet interventions?	Internet intervention methods: real-time video therapy, digital self-help programs, combining Internet interventions with face-to-face therapy, etc. Potential applications of AI in Internet interventions: virtual psychological coaches, psychotherapists, etc.	AI can work with therapists to improve outcomes.
(Blyler & Seligman, 2023a)	How a person's narrative identity can be used to provide coaches and therapists with a personalized approach to intervention.	ChatGPT-4 generates personalized narrative identities based on stream-of-consciousness thoughts and demographic information. Then, it provided targeted coaching methods and interventions based on narratives and analyzed the feasibility of the coaching strategies and specific interventions.	ChatGPT-4 generates highly credible coaching strategies and highly credible specific interventions based on the narrative identities it constructs.
Blyler and Seligman (2023b)	The potential of AI in psychological practice, AI-generated personal narratives in therapy and counseling to promote self-	Stream-of-consciousness reflections and basic demographic information were processed through the ChatGPT-4 to generate personal narratives and then evaluated these AI-generated narratives for accuracy, level of surprise, and illuminating.	AI can support self-discovery in psychotherapy and coaching.

	discovery.	
Abd-Alrazaq et al. (2019)	How the features and applications of chatbots are meeting the needs of people in the field of mental health.	Chatbots in mental health are characterized by a wide range of applications, a variety of interactions, rule-driven or machine-learning generation of responses, the use of virtual representatives, and stand-alone software or web-based platform implementations.
Elyoseph and Levkovich (2023)	This paper seeks to address the potential and limitations of the AI language model ChatGPT for suicide risk assessment.	ChatGPT underestimated the risk of suicide attempts in all scenarios, compared to mental health professionals.
Kjell et al. (2023)	This paper seeks to explore how the use of LLMs can change psychological assessment.	Analyze natural language responses using LLMs to extract mental health-related information and the strengths (e.g., accuracy, scope, parsing, and openness) and limitations (e.g., bias, risk, and ethical issues) of LLMs for assessing mental health.
Pal et al. (2023)	Bias in LLMs age and gender dimensions.	LLMs have the potential to transform psychological assessments away from reliance on rating scales and towards using the language in which people naturally communicate.
(J. M. Liu et al., 2023)	How to provide effective counseling services in the field of mental health support through the use of LLM.	By creating population subgroups based on age and gender, found that most of the cross-cutting subgroups exhibited amplification of bias.
Schueller and Morris (2023)	The use of LLMs in the field of clinical and counseling psychology and how they can provide better psychological interventions.	ChatCounselor is able to generate interactive and meaningful responses to provide personalized mental health support to users.
Sharma et al. (2023)	How AI systems can be utilized to assist peer supporters in improving empathy expression.	LLMs have a wide range of applications in clinical science and practice that can help humans provide better interventions, but will not completely replace therapists.
Graber-Stiehl (2023)	Are AI treatments ready for mainstream adoption in today's world?	AI can help peer supporters demonstrate higher levels of empathy.
Stade et al. (2023)	How to responsibly develop and evaluate LLMs while realizing their potential value in behavioral health.	Despite the promise of AI in mental health, there are still many ethical and safety issues with current AI therapies.
Zhong et al. (2023)	Challenges of using LLMs in psychiatric research and practice.	LLMs have great potential in the field of psychotherapy.
		LLMs have great potential in psychiatric research and practice, but there are concerns (e.g., reliability, accuracy, transparency, accountability, and ethical issues).

#### 4. LLMs in educational and developmental psychology

Within multilevel time scales of human behavior (Newell, 1990), educational and developmental psychology is primarily positioned at the relatively medium- to long-term level (see Fig.1), which reflects the ongoing learning and development that characterizes the educational process. Educational and developmental psychology is concerned with the learning process, the accumulation of knowledge, the development of skills, and the changes in the individual psyche within the educational environment. According to a national survey, it was found that only three months after the public release of ChatGPT, 40% of U.S. teachers used it weekly for lesson planning (Demszky et al., 2023).

Table 3 summarizes the applications of LLMs to educational and developmental psychology. The potential for using LLMs in educational and developmental psychology is manifold. They can facilitate personalized learning and play an important role in emotion recognition, mental health aids, educational assessment, and improving learning motivation. Specifically, LLMs learn from massive amounts of data from the Internet and books (Binz & Schulz, 2023b), can be used as more knowledgeable learning aids (Stojanov, 2023), provide personalized learning experiences (Kasneci et al., 2023), will enhance motivation to learn (Ali et al., 2023). For example, Stojanov (2023) explored the potential of ChatGPT as a learning tool using the following method: He began his learning journey by setting learning objectives and conversing with ChatGPT about its functionality over four hours. Over the next three hours, he continued the discussion with ChatGPT and watched some relevant videos on YouTube. He felt positive feedback from his interactions with ChatGPT and found it a motivating and relevant learning experience.

**Table 3** Applications of LLMs in educational and developmental psychology study.

Author	Research question	Research method	Key finding
Frank (2023)	LLMs show impressive capabilities, but it is not clear what abstractions underlie their behavior.	Ensure LLMs have not been pre-exposed to the stimuli used in the experiment. Selecting simplified stimuli. Need for evidence of convergence across multiple experimental tasks and developmental processes.	By drawing on approaches from developmental psychology, researchers can better understand the representations used by LLMs.
Han (2023)	LLMs can aid research in ethics education and development, especially involving empirical and practical inquiry.	Using the Behavioral Definitional Issue Test (bDIT). Examine ChatGPT's learning and reasoning capabilities by requesting it to read and extract moral information from the letter from Martin Luther King. Using three stories of moral exemplars, analyze whether ChatGPT produces emotional and motivational responses when observing these stories.	LLMs can serve as a useful tool in moral education and developmental research to predict the potential effects of educational inputs on students' cognitive, motivational, and behavioral processes.
Stojanov (2023)	Explore the experience of using ChatGPT as a more knowledgeable other to aid in the learning process, as well as the	Using an autobiographical research methodology, this study explored the role of ChatGPT as a support for those with more knowledge in the learning process, particularly in terms of understanding ChatGPT's technology.	Instant answers help to create a 'flow' experience, but in a state of 'immersion', users may overestimate their knowledge and understanding.

potential and limitations.			
Kasneci et al. (2023)	Opportunities and challenges of LLMs in education.	LLMs can help create and design educational content, provide personalized learning experiences, aid in language learning, research, and writing, conduct assessments and grading, and facilitate professional development.	LLMs have great potential in education and can provide many benefits to students and teachers (e.g., personalizing instruction, increasing student engagement and interaction, and creating educational content for learners with diverse needs).
Cai et al. (2023)	Explore whether the ChatGPT is similar to humans in language comprehension and production, and how it performs on multiple psycholinguistic tasks.	12 pre-registered psycholinguistic experiments: Speech: sound-shape association. Speech: sound-gender association. Words: word length and predictivity. Words: word meaning priming. Syntax: structural priming. Syntax: syntactic ambiguity resolution. Meaning: implausible sentence interpretation. Meaning: semantic illusions. Discourse: implicit causality. Discourse: drawing inferences. Interlocutor sensitivity: word meaning access. Interlocutor sensitivity: lexical retrieval. Interlocutor sensitivity: word meaning access.	ChatGPT shares similarities with humans in language comprehension and production but also shows different patterns in some areas.
Ali et al. (2023)	Explore the impact of ChatGPT on English language students' motivation from teachers' and students' perspectives.	A five-point Likert scale was used to collect information about participants' perceptions of the impact of the ChatGPT on learning English, as well as whether or not the ChatGPT increased the students' interest in English language learning, self-directed learning, interacting with others, and the enjoyment and pleasure of learning English.	ChatGPT positively impacts the motivation of English language learners.
Kosinski (2023)	Explores whether LLMs spontaneously generate Theoretical Minds.	The researchers used two types of gold-standard false-belief tasks: the Unexpected Contents task, also known as the Smarties task, and the Unexpected Transfer task, also known as the Maxi task or Sally-Anne test.	ToM, previously thought to be a uniquely human ability, may have emerged spontaneously as a byproduct of LLMs' improved language skills.

## 5. LLMs in social and cultural psychology

In time scales of human behavior (Newell, 1990), social and cultural psychology covers a predominantly long-term dimension (see Fig.1), reflecting its focus not only on social interactions but also on the long-term behavioral patterns and mental processes of individuals in their social environments. Social and cultural psychology studies how individual behavior is influenced by the social and cultural environment and others and how individuals affect the social and cultural environment and others. These studies usually focus on interpersonal interactions (Tajfel, 1982), group behavior, attitude formation and change, and social cognition. LLMs can simulate human responses and behaviors and be used to test theories and hypotheses of human behavior (Grossmann et al., 2023). In social and cultural psychology, LLMs can revolutionize the field by analyzing large amounts of textual data, modeling social interactions, and providing valuable insights into human behavior and social dynamics (Salah et al., 2023).

First, LLMs share many similarities with humans regarding social cognition. For example, research has found that LLMs have a variety of typical human cognitive biases in judgment and decision-making, such

as the anchoring effect, the representativeness heuristic, and the base rate neglect (Talboy & Fuller, 2023). In addition, cultural psychology research has shown that there are significant differences in the cognitive processes of Easterners and Westerners when processing information and making judgments (Nisbett et al., 2001) and that LLM consistently favors an Eastern holistic way of thinking (Jin et al., 2023).

Second, LLMs have also been shown to characterize human groups in social interaction settings. For example, it has been shown that LLMs replicate the results of Milgram's electroshock experiments (Aher et al., 2023), show better gaming abilities in specific games (Akata et al., 2023), and exhibit different risk-taking and pro-social behaviors under different emotional states (Yukun et al., 2023).

Next, LLMs can also serve as specific social and cultural psychological research samples. For example, one study explored the potential of LLMs to serve as valid proxies for specific human subgroups in social science research and found that LLMs contain information that goes far beyond superficial similarity, reflecting the complex interplay between ideas, attitudes, and sociocultural contexts that characterize human attitudes (Argyle et al., 2022). In addition, one study has tested LLM for personality and values, and the results show that their scores are all similar to those of human samples (Miotto et al., 2022).

Therefore, LLMs have many applications in social and cultural psychology, allowing one to test theories and hypotheses about human behavior in social and cultural interaction settings. For example, one study explores whether an AI chatbot can adapt its financial decisions and pro-social behaviors through emotional cues as humans do (Yukun et al., 2023). The experimental design is divided into two parts: Study 1, investment decision-making, was chosen as the topic for this study because human investment decisions are susceptible to emotional cues. It is hypothesized that the investment risk-taking behavior of AI chatbots will be lower than that of the control group when they receive fear emotional cues and higher than that of the control group when they receive joy emotional cues. By providing the bots with different emotional cues (fear, joy, or no emotion), their responses in terms of investment decisions were collected and analyzed. Study 2 measured the pro-social responses exhibited by an AI chatbot by providing it with emotional cues of anxiety and joy by donating money to a sick friend. Like Study 1, whether emotional cues influenced its pro-social behavior was explored by collecting and analyzing the robot's responses under different emotional cues. Table 4 summarizes the applications of LLMs to social and cultural psychology.

**Table 4** Applications of LLMs in social and cultural psychology study.

Author	Research question	Research method	Key finding
Atari et al. (2023)	LLMs have made great strides in generating and	Using World Values Survey (WVS) data: place LLMs on the spectrum of contemporary human	LLMs performed as outliers on psychometrics compared

Jin et al. (2023)	analyzing textual data, but how similar are they to different kinds of humans?	psychological change. Standard Cognitive Tasks: through multiple standard cognitive tasks. Thinking styles: comparing GPT responses to extensive cross-cultural data from 31 human groups. Self-concept: using an established self-concept task.	to large-scale cross-cultural data.
Schaaff et al. (2023)	Explore ChatGPT's way of thinking within the framework of cultural psychology, i.e., whether it tends to think holistically or analytically.	Two scales that directly measure cognitive processes: the Analytic-Holistic Scale (AHS) and the Ternary Categorization Task (TCT). In addition, two scales investigating differences in cultural thinking values were used: the Dialectical Self Scale (DSS) and the Self-Construction Scale (SCS).	In terms of cognitive processes, ChatGPT favors Eastern holistic thinking. However, in terms of value judgment, ChatGPT did not significantly favor the East or the West.
Salah et al. (2023)	The main research question of this thesis was to explore the empathetic abilities of ChatGPT.	Understanding and expressing emotions: ChatGPT generates responses based on prompts and compares them to the expected emotion categories. Parallel Emotional Responses: analyzing the emotional responses generated by ChatGPT when prompted with different emotional categories. ChatGPT's empathetic personality: five psychologically-recognized questionnaires (e.g., IRI, EQ, TEQ, PES, and AQ) were used for a system-level assessment of ChatGPT's level of empathy in different areas.	ChatGPT can understand the emotions of others and take their perspective but still has some difficulty in showing higher levels of empathy compared to healthy humans.
Harding et al. (2023)	Explore the use of generative AI (e.g., ChatGPT) in social psychology research, including its potential advantages and limitations.	Simulating social interactions: providing insights into human behavior and social phenomena. Analyzing large amounts of textual data: providing an in-depth understanding of human behavior and social interaction. Extracting insights about cognitive processes: providing an in-depth understanding of the role of cognitive processes in social behavior and social interaction.	ChatGPT has great potential in social psychology research to help analyze large amounts of textual data, simulate social interactions as well as provide valuable insights into human behavior and social dynamics.
Patel and Fan (2023)	Can LLMs replace human research participants, especially in the field of moral psychology research?	LLMs replace human participants in moral psychology: to help generate and refine research hypotheses, to pilot test items, and to validate data from human subjects.	Despite the potential of LLMs to simulate human behavior and thinking, it is unlikely that they can fully replace humans as participants in scientific research.
X. Wang et al. (2023)	The ability of three current language models (Bard, GPT 3.5, and GPT 4) to recognize and describe emotions, as well as their levels of empathy.	Assessing Emotional Intelligence in LLMs by Developing a Standardized Emotional Understanding Test (SECEU): The test is based on scenarios designed to elicit positive and negative emotions based on school, family, and social contexts. Participants were asked to assign 4 possible emotions to each scenario (e.g., surprise, joy, confusion, pride) a total rating of 10.	Current LLMs have human-equivalent capabilities in sentiment recognition and description.
Hutson (2023)	Explore whether LLMs (e.g., GPT-3.5) can replace human participants and their potential applications in some fields.	Ethical judgment experiment: use GPT-3.5 to assess 464 ethical scenarios that had been previously assessed by human participants. Consumer Behavior Experiment: use GPT-3.5 to simulate consumer behavior to test its ability to replace human participants in market research. Diversity of AI Participants: allow GPT-3 to exhibit different personality types by providing it with different character traits. Social Simulation Experiment: use GPT-3 to create a virtual social platform with 1,000 different users called SimReddit.	LLMs varied widely in their performance on Emotional Intelligence (EI).
Grossmann et al. (2023)	How to revisit and improve social science research methods in the context of the ongoing impact of AI, especially LLMs, on social science research.	Potential applications of LLMs in a variety of social science research methods such as questionnaires, behavioral tests, mixed methods analysis of semi-structured responses, agent-based modeling (ABM), observational studies, and experiments.	LLMs can be used in place of human participants for data collection, and also can be used in agent-based modeling (ABM) to explore how individuals with specific characteristics and beliefs influence subsequent human

		interactions.
Akata et al. (2023)	How LLMs interact with other LLMs and simple human-like strategies in repeated games, and how their behavioral characteristics are reflected in different types of games.	Getting LLMs to play limited repetition games: analyze their behavior when playing against other LLMs as well as simple human-like strategies. Analyzing behavior in different game families: Prisoner's Dilemma and War and Gender games.
Abramski et al. (2023)	Analyzing and revealing the biases of LLMs such as GPT-3, GPT-3.5, and GPT-4 in describing math and STEM disciplines.	Construction of Behavioral Formal Mental Networks (BFMNs). Semantic frame analysis: analyze the semantic frames generated by LLMs to understand how they relate mathematics to other concepts. Examining different versions of LLMs: comparing GPT-3, GPT-3.5, and GPT-4 perceptions and biases in math and STEM domains. Compare with high school student data.
Yukun et al. (2023)	Whether LLMs can adjust their financial decisions and pro-social behaviors in response to emotional cues, and whether this ability increases with advances in language models.	Collecting and analyzing the responses of the robot in terms of investment decisions by providing it with different emotional cues (fear, joy, or no emotion). Measuring the pro-social responses exhibited by an AI chatbot through donations to a sick friend by providing it with emotional cues of anxiety and joy.
Suri et al. (2023)	Whether LLMs exhibit decision-making heuristics and biases similar to humans, and whether LLMs influence human cognition and decision-making processes.	Anchoring Heuristic: Create a low anchor and a high anchor prompt and ask ChatGPT to estimate the books' numbers. Representativeness and Availability Heuristic: A structure similar to Linda's question was created to test whether ChatGPT would choose more typical options. Framing effects approach: two scenario prompts (a positive and a negative frame), ChatGPT was asked to rate the efficacy of the drug. End-beat effect: ask ChatGPT to choose one of two identical coins to donate to a museum.
Talboy and Fuller (2023)	How do human cognitive biases permeate the output of LLMs?	Representativeness heuristic: asking ChatGPT and Bard which career a person with a particular characteristic would be most likely to pursue. Base rate neglect and value selection bias: using a well-known base rate neglect prompt. Anchoring and adjustment: gave a random number and asked the model to indicate whether this number was higher or lower than the UN rate for African countries. Framing effects: using an established framing paradigm, a hypothetical disease and its two potential treatment options.
Park et al. (2022)	How to generate a prototype with real social interaction to help designers understand and adjust to the challenges that can arise when social computing systems are filled at scale.	Test whether social behaviors generated by a social simulator are credible across multiple never-before-seen communities. Generate 50 discussions from GPT-3 for a subreddit created after November 2020. Randomly select an actual discussion and a generated discussion, and have participants identify which one is real.
X. Li et al. (2022)	To assess whether LLMs are safe on a psychological level and whether they exhibit features suggestive of mental illness.	Selection of LLMs: LLMs, GPT-3, InstructGPT and FLAN-T5-XXL. Psychological tests: Two personality tests (SD-3 and BFI) and two well-being tests (FS and SWLS) were used to assess the LLMs. Assessment Framework: Unbiased prompts were designed. For each prompt and statement, three outputs were sampled from the LLM and mean scores were calculated.
Sap et al. (2022)	Whether LLMs have social intelligence and Theory of Mind (TOM), and how to improve these abilities.	The SOCIALIQA dataset was used to assess the social intelligence and general knowledge of LLMs. The TOMI dataset was used to test the ability of LLMs to reason about the mental states and reality of others.

Miotto et al. (2022)	Examine personality traits, held values, and self-reported demographic characteristics of the GPT-3 by assessing them with two validated measurement instruments.	Personality (HEXACO scale) and values (Human Values Scale) of GPT-3 were assessed using two validated measurement instruments. Experiments were run with different temperature parameter settings to assess whether temperature affects GPT-3's personality and values. Data were analyzed to report personality traits, values, and demographic characteristics of GPT-3 at different temperatures and compared to a human baseline study.	social intelligence and theory of mind.
Argyle et al. (2022)	Can LLMs serve as effective proxies for modeling the performance of specific human subgroups in social science research.	Use GPT-3 to generate a story context about Democrats and Republicans for each person in the survey, and then ask about the vocabulary newly sampled by GPT-3. Use GPT-3 to generate in silico samples for 2012, 2016, and 2020 ANES participants that match their demographic characteristics and compare them to the corresponding human samples. Use GPT-3 to generate associations of complex patterns about a wide range of conceptual nodes to assess its algorithmic fidelity in complex structural associations.	The "algorithmic bias" in GPT-3 is fine-grained and demographically relevant, which can enable it to accurately model the distribution of responses from various human subgroups.
Trott et al. (2023)	Can LLMs understand other people's beliefs as well as humans do?	GPT-3 were tested for their performance on a written version of the False Belief Task (FBT) to assess their sensitivity to a character's viewpoint state in a written passage. By comparing the performance of human participants and GPT-3 on this task, we to understand whether linguistic input is sufficient to account for the ability of humans to reason about others' mental states.	GPT-3 still performs less well than humans and does not fully explain human behavior when dealing with false belief tasks.
Aher et al. (2023)	How to simulate multiple human behaviors using LLMs and to evaluate the accuracy and consistency of these models in simulating specific human behaviors.	Ultimatum Game: to evaluate the accuracy of LLMs in simulating human behavior by generating random completions. Garden Path Sentences (GPS): to assess the performance of models in psycholinguistic research. Milgram Shock Experiment (MSE): to assess the performance of different LLMs in social psychology research. Wisdom of Crowds: to assess the performance of different LLMs in interdisciplinary research.	In the experiment (Wisdom of Crowds), LLMs exhibited an "over-accuracy distortion", which may affect downstream applications, such as education and the arts.

## 6. LLMs as research tools in psychology

The LLMs are a powerful tool for scientific research and can be used as a research aid to help psychologists with everything from literature review, hypothesis generation, experimental design, experimental subjects, and data analysis to academic writing and peer review (see Table 5)

**Table 5** LLMs as research tools in psychology study.

Topic	Related study
Literature review	LLMs can summarize the researched literature (Dis, Bollen, Zuidema, Rooij, & Bockting, 2023), complete literature review tasks (Qureshi et al., 2023), and create literature review articles (Aydin & Karaarslan, 2022), at the same time, there are LLM that has been specially trained to accomplish systematic literature reviews (Taylor et al., 2022).
Hypothesis generation	LLMs can generate hypotheses from scientific literature, make inferences based on scientific data, and then clarify their conclusions through interpretation (Zheng et al., 2023), and can quickly and automatically test these research hypotheses and learn from mistakes (Park et al., 2023).
Experimental design	LLMs provide text-based material for experimental design, thereby optimizing the research process and reducing experimental complexity. By employing these models, researchers can easily create experimental stimuli, develop test items, and even simulate interactive sessions in controlled

---

	environments (Aher, Arriaga, & Kalai, 2022; Akata et al., 2023), providing a high degree of control and precision to the experimental process.
Experimental subjects	LLMs can simulate some human behaviors and responses, which provides an opportunity to test theories and hypotheses about human behavior (Grossmann et al., 2023), their use in place of human participation in experiments saves time and costs and can be applied to some experiments where human participation is not appropriate (Hutson, 2023), they can be combined with factors such as the specific research topic, the task, and the sample, and the use of LLM as an alternative to research participants where appropriate (Dillion et al., 2023).
Data analysis	LLMs can efficiently analyze massive amounts of textual data to gain insights into human behavior and emotions at an unprecedented scale (Patel & Fan, 2023), can analyze textual data in multiple languages, and accurately detect mental structures within it (Rathje et al., 2023), can draw mental profiles from social media data (Peters & Matz, 2023).
Academic writing	LLMs can also help humans in writing (Dergaa et al., 2023; Stokel-Walker, 2022; Van Dis et al., 2023).
Peer review	LLMs were used in two natural language processing tasks and a human expert to assess the quality of the text, and the results of the assessment were consistent with those of the human expert (Chiang & Lee, 2023), LLMs offer the opportunity to get things done quickly, from Ph.D. students struggling to finish their dissertations, to peer reviewers submitting analyses under time pressure(Van Dis et al., 2023).

---

## 6.1. Automated literature review and meta-analysis

Conducting a literature review meta-analysis is a complex, arduous process that requires significant expertise and time (Michelson & Reuter, 2019). Nature reports that researchers have used ChatGPT as a research assistant to summarize the literature of studies (Dis, Bollen, Zuidema, Rooij, & Bockting, 2023). In one study, researchers utilized ChatGPT to complete some systematic literature review tasks (Qureshi et al., 2023). In another study, a literature review article was created using ChatGPT with the application of digital twins in the health field as the theme, and the results showed that knowledge compilation and representation were accelerated with the help of ChatGPT. However, academic validity needs to be further verified (Aydin & Karaarslan, 2022). Meanwhile, there are also LLMs specifically trained by researchers for the practical needs of scientific research (Taylor et al., 2022), which can accomplish a systematic literature review.

In summary, LLMs speed up the process of literature review and meta-analysis. Researchers can use these models to systematically review and synthesize existing research, improving the efficiency of evidence-based psychology.

## 6.2. Hypothesis generation and experimental design

Hypothesis-driven research is at the core of scientific activity. LLMs can generate hypotheses from scientific literature, make inferences based on scientific data, and then clarify their conclusions through interpretation (Zheng et al., 2023). Although LLMs are capable of generating research hypotheses and becoming better "hypothesis machines," they will need to improve their logical and mathematical derivation capabilities in the future to eliminate more factual errors and be able to quickly and automatically test these research hypotheses to learn from their mistakes (Y. Park et al., 2023). As an innovative tool, LLMs have great potential for use in psychological experiments. They can provide text-based material for experimental

design, optimizing the research process and reducing experimental complexity. By employing these models, researchers can easily create experimental stimuli, develop test items, and even simulate interactive sessions in controlled environments (Aher, Arriaga, & Kalai, 2022; Akata et al., 2023), providing a high degree of control and precision to the experimental process.

In conclusion, LLMs provide a powerful and flexible tool for psychological research, from hypothesis generation to experimental design, to help researchers achieve more efficient and precise research goals.

### **6.3. As subjects in a psychological experiment**

Although LLMs can simulate some human behaviors and responses, which provides an opportunity to test theories and hypotheses about human behavior (Grossmann et al., 2023), there is still some controversy as to whether LLMs can be used as a substitute for human subjects to participate in psychological research. Some researchers have argued that LLMs can be used in psychology as a substitute for human participation in experiments to save time and cost and can be applied to experiments that are not suitable for human participation while recognizing that these models can have some problems (e.g., bias and insufficiently trained data, etc.) (Hutson, 2023). Some other researchers have proposed using LLMs as an alternative method of studying participants when appropriate, based on their performance in conjunction with factors such as specific research topics, tasks, and samples (Dillion et al., 2023). Some researchers believe that although LLMs will significantly impact scientific research, they are unlikely to replace human participants in any meaningful way (Harding et al., 2023). Although there is some controversy about whether LLMs can replace humans as experimental subjects, some studies of LLMs as subjects have shown that LLMs perform similarly to humans (Orrù, Piarulli, Conversano, & Gemignani, 2023; Peter S. Park, Schoenegger, & Zhu, 2023), which may indicate the potential of LLMs to replace humans as subjects.

In conclusion, although LLMs can simulate human judgments, their understanding of human thinking is still limited, and their output should be validated and interpreted with caution when chosen as psychological subjects.

### **6.4. Tools for data analysis**

Various forms of AI have long been used to analyze psychological data, such as flight data for pilot screening (Ke et al., 2023). Machine learning algorithms have facilitated the processing of large datasets, identifying patterns and correlations that may have been overlooked. However, LLMs take this capability to a new level. These models can efficiently analyze massive amounts of textual data to gain insights into

human behavior and emotions on an unprecedented scale (Patel & Fan, 2023). Psychological research means faster and more comprehensive data analysis, leading to more reliable and nuanced findings. LLMs can analyze textual data in multiple languages, accurately detect psychological structures within them (Rathje et al., 2023), and go for psychological profiles from social media data (Peters & Matz, 2023). In addition, LLMs have demonstrated a degree of competence in the medical field; LLMs can predict the optimal neuroradiographic imaging modality for a given clinical presentation, even though LLMs do not outperform experienced neuroradiologists, suggesting the need for continued improvement in the medical context (Nazario-Johnson et al., 2023). These findings demonstrate the great potential of LLMs in evaluating and analyzing data.

## 6.5. Paper writing and peer review tools

It has been argued that LLMs are not currently a complete replacement for human writing, but instead answer questions and generate naturally fluent and informative content compellingly, but with no real intelligence, just generated text based on patterns of previously seen words (Stokel-Walker, 2022). In one study, students used ChatGPT as an aid in their writing. However, the results showed that the experimental group that used ChatGPT was similar to the control group in terms of writing quality, speed, and authenticity, and the authors suggest that this may be because experienced researchers can better guide ChatGPT to high-quality information. In contrast, the students may be having ChatGPT difficulties (Bašić, Banovac, Kružić, & Jerković, 2023). In another article, the authors discuss the prospects and potential threats of ChatGPT in academic writing and emphasize that using ChatGPT in academic research should prioritize peer-reviewed scholarly sources. In addition, the article mentions the potential advantages of ChatGPT in academic research, including the handling of large amounts of textual data, automatic generation of abstracts, and research questions (Dergaa, Chamari, Zmijewski, & Saad, 2023). Additionally, LLMs have the potential for peer review (Van Dis et al., 2023), where the results of LLM's evaluation in a text evaluation task are consistent with those of human experts (Chiang & Lee, 2023).

In conclusion, LLMs such as ChatGPT are potent tools for academic writing, capable of processing large amounts of textual data and automating tasks that were previously done manually; it can be used to scan academic papers and extract essential details, generate objective and unbiased abstracts, and create research questions. It also has the potential to be applied to peer review of papers. However, researchers must exercise caution when using them as they can also integrate false or biased information into papers,

leading to unintentional plagiarism and misattribution of concepts (Dis, Bollen, Zuidema, Rooij, & Bockting, 2023).

## 7. Challenges and future directions

### 7.1. Challenges and limitations

Although the potential of LLMs to simulate complex cognitive processes is enormous, providing researchers with new tools to explore the mechanisms of human cognition and behavior opportunities for a wide range of applications in a variety of fields, including clinical and counseling, educational and developmental, and social and cultural psychology. However, the output of LLMs should not be mistaken for the presence of thoughts but instead be viewed as complex pattern matching based on probabilistic modeling (Floridi & Chiriatti, 2020). Although its performance is impressive, this is different from the model being conscious or genuinely understanding. The interpretation of its capabilities must be based on understanding its limitations and the nature of its operation, which may differ fundamentally from human cognition. Therefore, it is essential to focus on the potential of LLMs in psychological research while at the same time facing up to the technical and ethical challenges that may arise.

First, despite the emergence of competence in the LLM (Wei et al., 2022), its internal working mechanism remains a black box from a cognitive and behavioral psychology perspective. For example, LLMs perform impressively on tasks requiring formal linguistic competence (including knowledge of the rules and patterns of a particular language) but fail many tests requiring functional competence (the set of cognitive abilities needed to understand and use language in the real world) (Mahowald et al., 2023), excels in analogical reasoning and moral reasoning tasks, but performs poorly on spatial reasoning tasks (Agrawal, 2023).

Second, while LLMs have accelerated the use of AI technology in clinical and counseling psychotherapy, privacy and ethical issues may arise (Graber-Stiehl, 2023). For example, it has been shown that gatekeepers, patients, and even mental health professionals who rely on ChatGPT to assess suicide risk or use it as an aid to improve decision-making may receive inaccurate assessments that underestimate actual suicide risk (Elyoseph & Levkovich, 2023), and may also bias clinician decision-making, which can lead to healthcare inequity (Pal et al., 2023). In addition, LLMs in psychiatry research and practice have been associated with potential bias and privacy violations (Zhong et al., 2023).

Third, in fields such as educational, developmental, and social and cultural psychology, LLMs face problems and challenges in their application. For example, when applied in education, LLMs have the potential for output bias and misuse (Kasneci et al., 2023). One study found that the texts generated by ChatGPT were not always consistent or logical and sometimes even contradictory (Stojanov, 2023). In the field of social and cultural psychology, LLMs exhibit cognitive biases (Talboy & Fuller, 2023) and cultural biases (Atari et al., 2023) similar to those of humans, in addition to implicitly darker personality patterns (X. Li et al., 2022). Field Bender et al. (2021) have argued that training data for LLMs may reflect social biases that continue to be perpetuated in research settings.

Finally, as an aid to scientific research, LLMs have some limitations. For example, when it comes to writing, LLMs currently do not fully replace humans. Instead, they answer questions and generate naturally flowing and informative content compellingly, without real intelligence, only generating text based on previously seen word patterns (Stokel-Walker, 2022). Although macrolanguage models can simulate human judgments when used as experimental subjects, there are still limits to their understanding of human thought (Dillion et al., 2023). Field Van Dis et al. (2023) noted that LLMs may accelerate innovation, shorten publication times, and increase scientific diversity and equality. However, they may also reduce the quality and transparency of research and fundamentally alter scientists' autonomy as human researchers.

In summary, while LLMs offer extraordinary capabilities for psychological research, they also present challenges related to bias, ethical issues, data security, transparency, and technical expertise. Researchers should be fully aware of these challenges when using big language models and take steps to address them responsibly in their research projects. The following table summarizes the challenges and limitations of LLMs in psychological applications (see Table 6).

**Table 6** Challenges and limitations of LLMs in psychological applications.

Author	Domain	Challenges and limitations
Mitchell (2023)	Cognition and Behavior	Lack of real-world understanding. Lack of abstract reasoning. Lack of understanding of user intent.
Stella et al. (2023)	Cognition and Behavior	Lack of meta-knowledge leads to some limitations of LLMs in processing information). Lack of curiosity and which raises questions about the source of the "creativity" they exhibit). Hallucinations: LLMs unconsciously fabricate information and are unable to identify the source of their knowledge.
Sartori and Orrù (2023)	Cognition and Behavior	Lack of causal reasoning ability: they may not perform well in causal reasoning. Dependence on training data: if the training data is biased, the model may not perform well in other tasks. Lack of creativity and imagination.
Goertzel (2023)	Cognition and Behavior	Lack of autonomy: LLMs are unable to systematically pursue complex goals. Lack of abstract reasoning: LLMs perform poorly in performing highly complex multi-step reasoning. Lack of self-understanding: LLMs are unable to reflect fully on their behavior and limitations. Lack of in-depth understanding of the real world: Leading to potential problems when they perform tasks involving the real world.

Peng et al. (2023)	Cognition and Behavior	Forgetting problem: LLMs may forget previously learned knowledge when learning new tasks. Inadequate common-sense reasoning: LLMs perform poorly on common sense reasoning tasks. Lack of systematic demonstration of problem-solving skills: studies have found that LLMs occasionally perform poorly when solving problems.
Holtzman et al. (2023)	Cognition and Behavior	Lack of clear understanding of model behavior: making it difficult to improve model performance and solve problems. Lack of formal description of model behavior: the lack of formal description of model behavior makes it difficult for researchers to systematically analyze model behavior and thus find a unified theory to explain model behavior. Lack of interpretability of model behavior: it difficult for researchers to understand how models can perform well on some tasks and poorly on others.
Seals and Shalin (2023)	Cognition and Behavior	ChatGPT and human-generated analogies differed in these stylistic dimensions, these lexical features, their choice of words for these features and these devices that help readers understand text. ChatGPT may lack human cognitive and psycholinguistic features when generating analogies.
Stade et al. (2023)	Clinic and Counseling	Technical limitations: may have difficulty in assessing patients for suicide risk, substance abuse, safety issues, medical comorbidities, and life events). Connecting with the patient: may have difficulty interpreting nonverbal behaviors. Problems with full autonomy and therapeutic relationship (e.g., altering the patient's existing relationships or social skills).
Li et al. (2023)	Education and Development	Academic integrity and the definition of authorship. Assessment methods and educational consequences. Data privacy and security. Teacher-student relationship. Students' critical thinking skills. Misinformation and bias. Interpersonal communication skills development.
Kasneci et al. (2023)	Education and Development	Technical limitations: leading to insufficient personalization and adaptation. Bias and equity: affecting teaching and learning processes and outcomes. Over-reliance on models: leading to a decline in creativity, critical thinking, and problem-solving skills.
		Inadequate knowledge and expertise: Many educators and institutions may lack the knowledge and expertise to effectively integrate new technologies. Maintenance costs. Multilingual support and equitable access.
Fecher et al. (2023)	Society and Culture	Liability issues: challenging traditional mechanisms of authorship and liability. Bias issues: affecting the objectivity and impartiality of science. Privacy and data protection issues: may be privacy issues with the training data of LLMs. Intellectual property issues: potential legal disputes. Environmental issues: generating large amounts of carbon emissions, which can have a negative impact on the environment.
Atari et al. (2023)	Society and Culture	Ignoring global psychological diversity (e.g., tend to favor the psychological characteristics of WEIRD societies) and which can lead to prejudice and discrimination against people of other cultures and backgrounds. Differences in values and moral judgments and which can lead to problems of communication and understanding in multicultural societies. Self-identity and perceived social roles and which may lead to stereotypes and misconceptions about non-WEIRD populations).
Park et al. (2023)	Society and Culture	Reduced innovation and development, bias and discrimination, culture clash and conflict, differences in values and morals and entrenchment of the status quo.
Salah et al. (2023)	Society and Culture	Limited understanding of social context: Although ChatGPT performs well in syntax and general semantics, it still has limitations in capturing the nuances of social language. Ethical challenges: AI-generated fake content can lead to ethical issues including digital personhood, informed consent, potential manipulation, and the implications of using AI to simulate human interactions.
Hayes (2023)	Society and Culture	Potential biases: if the training data contain biases, LLMs may learn and replicate them. Data privacy and consent issues: Text generated using LLMs may involve data privacy and consent issues. Output may be non-humanly understandable: although LLMs generate text that closely resembles human language, they do not truly understand the content and may generate absurd or misleading responses.
Miotto et al. (2022)	Society and Culture	Bias and discrimination: LLMs may be affected by biases in the training data, which can produce unfair results, such as reinforcing sexism in the translation of job advertisements. Responsibility and control: Due to the complexity of language models, it is difficult to determine who is responsible for the model's output, which can lead to attribution of problems and lack of controls.
Bender et al. (2021)	Society and Culture	Potential Harm: LLMs may lead to the propagation of harmful ideas such as stereotyping, discrimination, and extremism, and may lead to misinformation and bullying when generating text. Data bias and unfairness: leading to potential harm to marginalized communities. Automating bias: exacerbating existing biases and discrimination. Enhancement of authoritative viewpoints: LLMs may reinforce dominant viewpoints in the training data, further undermining marginalized people.
Tamkin et al. (2021)	Society and Culture	Alignment: In order to better align models with human values, algorithmic improvements are needed to increase factual accuracy and robustness against adversarial samples. In addition, appropriate values need to be made explicit for different usage scenarios. Societal Impact: Widespread use of LLMs may lead to problems such as information leakage and amplification of bias.
Brown et al. (2020)	Society and Culture	Misuse of language modeling: GPT-3 may be used to generate fake news, spread extremist ideas, conduct cyber-attacks and other malicious uses. Fairness, bias, and representation: GPT-3 may carry bias against gender, race, and religion, among others, sparking related controversies. News generation: News generated by GPT-3 may be difficult to distinguish

		from real news, leading to confusing and misleading information.
Sallam (2023)	Research tools	Plagiarism: content generated by ChatGPT may be considered plagiarized, violating academic norms. Copyright issues: Is the generated content owned by ChatGPT or by the user? Transparency issues: The workings of ChatGPT may not be transparent, making it difficult for users to understand the source of generated content. Liability issues: who is responsible for ChatGPT when generating incorrect content?
Gupta et al. (2023)	Research tools	Transparency and Explanation: The working mechanism of generative AI models may be difficult to explain, which may lead users to doubt the credibility of the generated content. Legal and Ethical Issues: Generative AI models may involve intellectual property, privacy, and ethical issues, requiring attention to compliance with relevant laws and regulations during use.
Dergaa et al. (2023)	Research tools	Integration of erroneous or biased information. Problems with citing original sources and authors. Impact on academic integrity and quality. Increased inequity and inequality: Difficulty in recognizing AI-generated content. Academic evaluation and recognition issues. Direct replacement for academic researchers: ChatGPT is not a complete replacement for academic researchers as it has limitations in certain types of academic research.
Peters and Matz (2023)	Research tools	User privacy: LLMs can infer psychological traits from a user's social media data, which may violate the user's privacy. Potential bias: LLMs may create potential bias in the inference process, which may lead to unfair treatment of specific groups (e.g., gender, age, etc.). Data security: if the inferential power of LLMs is used maliciously, it may lead to data leakage, with serious implications for users' mental health.
Y. Liu et al. (2023)	Research tools	Academic misconduct: ChatGPT may be used for academic cheating, such as generating false papers or assignments. Challenges in the medical field: ChatGPT has limitations in medical image analysis, which may lead to wrong diagnosis and jeopardize patients' health.

## 7.2. Future directions and emergent trends

The LLMs have begun to be used in different areas of psychology, especially in cognitive and behavioral, clinical and counseling, educational and developmental, and social and cultural psychology. As the capabilities of LLMs are further enhanced, their applications in psychology still have the potential to continue to develop in the future.

First, in the field of cognitive and behavioral psychology, with the emergence of multimodal LLMs (OpenAI, 2023), on the one hand, it is possible to combine visual and auditory information with textual data to understand better and model emotions, behaviors, and mental states for cognition, on the other hand, it is possible to use neuroimaging data to inform the architectures and parameters of LLMs, and to integrate this information with traditional textual data integration to create more accurate and biologically sound models of human language and thought.

Second, in the field of clinical and counseling psychology, on the one hand, personal data such as social media posts, medical records, or wearable device data can be used to create tailored and personalized LLMs that provide more accurate and relevant insights into an individual's state of mind, on the other hand, the strengths of combining the clinical and counseling expertise of the human psyche with the scalability and computational power of an LLM can be combined to create new diagnostic treatment and intervention tools. In addition, in the fields of educational and developmental, and social and cultural psychology, it is essential

to build ethical LLMs and to ensure that they are designed and deployed in a way that respects privacy and uses data fairly and responsibly.

Ultimately, LLMs are a systematic project whose future development cannot be achieved without the interdisciplinary collaboration of researchers in fields as diverse as psychology, computer science, and linguistics, and for psychology researchers, an accessible open-source large language modeling framework and tools may be an integral part of future research efforts. The following table summarizes LLMs' future directions and emergent trends in psychological applications (see Table 7).

**Table 7** Future directions and emergent trends of LLMs in psychological applications.

Author	Domain	Future directions and emergent trends
D'Oria (2023)	Cognition and Behavior	Delving into Human-Computer Interaction (HCI) to understand AI's ability to mimic human behavior. Exploring how AI language modeling can be applied in the human sciences to improve research efficiency and quality
Crockett and Messeri (2023)	Cognition and Behavior	Focus on the costs of adopting alternative human narratives in cognitive science research, such as masking the human labor behind them and the impact on human well-being. Concern about the impact of technological developments on scientific work and human understanding to ensure that cognitive scientists remain proactive in technological advances.
Binz and Schulz (2023b)	Cognition and Behavior	Explore ways to make LLMs more stable and robust in the face of descriptive tasks. Investigate whether LLMs can learn to explore purposefully and how to better utilize causal knowledge in tasks. Analyze the performance of LLMs in different tasks and contexts to see if they can adapt like humans. Explore how LLMs develop and refine their cognitive abilities during natural interactions with humans.
Huang and Chang (2022)	Cognition and Behavior	Improve the reasoning ability of LLMs to encourage reasoning by optimizing training data, model architecture, and optimization goals. Develop more appropriate evaluation methods and benchmarks to measure the reasoning ability of LLMs to better reflect the true reasoning ability of the models. Investigate the potential of LLMs in different applications (e.g., problem solving, decision making and planning tasks). Explore other forms of reasoning (e.g., inductive and retrospective reasoning).
Abd-Alrazaq et al. (2019)	Clinic and Counseling	Develop more chatbots for people with mental illness, especially for those with disorders such as schizophrenia, obsessive-compulsive disorder and bipolar disorder.
Stade et al. (2023)	Clinic and Counseling	Implement more chatbots in developing countries to address the shortage of mental health professionals. Conduct more randomized controlled trials to evaluate the effectiveness of chatbots in mental health.
Demszky et al. (2023)	Clinic and Counseling	Developing new therapeutic techniques and evidence-based practices (EBPs). Focus on evidence-based practices first: to create meaningful clinical impact in the short term, clinical LLM applications based on existing evidence-based psychotherapies and techniques will have the greatest chance of success. Involve interdisciplinary collaboration. Focuses on therapist and patient trust and usability. Criteria for designing effective clinical LLMs.
Hagendorff (2023)	Education and Development	Developmental psychology: examining how LLMs develop cognitively, socially, and emotionally over the lifespan and how these models can be optimized for specific tasks and situations. Learning psychology: studying how LLMs acquire and retain knowledge and skills, and how to optimize these models to improve learning.
Sap et al. (2022)	Society and Culture	Explore more interactive and empirical training methods to help LLMs acquire true social intelligence and theoretical mental abilities. Investigate ways to combine static text with rich social intelligence and interaction data to improve social intelligence in LLMs. Investigate the theoretical-psychological abilities of LLMs in more naturalistic settings to reveal their performance in real-world scenarios.
Argyle et al. (2022)	Society and Culture	Investigate the algorithmic fidelity of the GPT-3 model and how appropriate conditioning can allow the model to accurately simulate the response distributions of various human

		subgroups. Created "in silico samples" by conditioning on the socio-demographic backgrounds of real human participants in multiple large U.S. surveys.
Schaaff et al. (2023)	Society and Culture	Developing more advanced models: to more accurately capture the emotional context of conversations and improve emotional understanding and expression. Measuring the emotional capabilities of bots: to investigate how to assess the emotional capabilities of chatbots in order to better understand how they behave when interacting with humans. Explore the use of ChatGPT as a support tool: investigate how ChatGPT can be used to support people more empathetically and improve human well-being.
Ziems et al. (2023)	Society and Culture	Cross-cultural CSS research: future research should separately consider the utility of LLMs for cross-cultural CSS in order to better serve social science research in different cultural contexts. Future research could explore contrastive or causal explanations in LLMs. New paradigms for social science and AI collaboration.
Van Dis et al. (2023)	Research tools	Invest in truly open LLMs: develop and implement open-source AI technologies to increase transparency and democratic control. Embrace the advantages of AI: utilize AI to accelerate innovation and breakthroughs at all academic stages, while focusing on issues of ethics and human autonomy. Broaden the discussion: organize international forums to discuss the development and responsible use of LLMs in research, including issues of diversity and inequality.
Fecher et al. (2023)	Research tools	Analyzing the risks and opportunities of LLMs for science systems. Examining how LLMs affect academic quality assurance mechanisms, academic misconduct, and scientific integrity. Exploring the impact of LLMs on academic reputation, evaluation systems, and knowledge dissemination. Examining how to balance the potential benefits from LLMs with adherence to scientific principles.

## 8. Conclusion

With the rapid development of AI technology, especially the continuous advancement of LLMs such as the GPT family, we have entered a new era characterized by an unprecedented level of machines able to understand and generate human language. This development is not just a technological breakthrough for the field of psychology but opens the door to a range of potential applications.

First, in the field of cognitive and behavioral psychology, LLMs are excelling in a variety of cognitive tasks. Although there are still limitations in causal cognition and planning, these models resurrect the principle of association, demonstrating the ability to associate at a distance and reason in complex ways. At the same time, the ability to adapt LLMs to cognitive models is a significant strength of psychological research, allowing new explorations of human cognitive and behavioral processing mechanisms.

Second, in clinical and counseling psychology, LLMs can be used as a preliminary diagnostic tool for mental health. While traditional mental health diagnosis relies on the experience of professionals and direct interaction with patients, LLMs can quickly identify potential mental health problems, such as depression and anxiety, by analyzing an individual's verbal expressions and textual content. Of course, such a diagnosis cannot wholly replace a professional psychological assessment. However, it can serve as an effective adjunct to help psychologists understand a patient's condition more quickly or play a role in primary mental health interventions. Meanwhile, personalized psychological intervention is another critical application direction

of the LLM. By combining information about an individual's health data and lifestyle habits, these models can provide tailored psychological advice and intervention programs. This personalized approach may be crucial for improving the effectiveness of psychological interventions.

Third, in educational and developmental and social and cultural psychology, LLMs have the same potential for application. For example, these models provide interactive and personalized learning experiences or generate research tasks based on real-life case applications that increase motivation and enhance learning. In addition, by analyzing large amounts of social media data, these models can help researchers track and analyze public sentiment changes to understand psychosocial dynamics better.

Finally, in psychological research, LLMs can drastically improve research efficiency. Researchers can use these models to quickly organize and analyze large amounts of literature, thus saving time. In addition, these models can also assist in experimental design, data analysis, and even writing papers, making psychological research more efficient and precise.

In summary, the applications of LLMs in psychology are promising. From research aids to cognitive modeling, from individualized interventions to personalized learning, and from the cognitive abilities of individuals to the social interactions of groups, these models have the potential to dramatically improve the understanding of the patterns of human communication, thought processes, and behaviors that lead to the development of more sophisticated theories of mind. However, despite the great potential for applying LLMs in psychology, being wary of the risks and challenges involved is essential. Ensuring these applications adhere to ethical standards is vital, especially in protecting individual privacy and data security. It is also important to realize that no matter how technologically advanced, LLMs can only partially replace the judgment and experience of human professionals. Therefore, these models should be viewed as an aid rather than an all-in-one solution.

## References

Abd-Alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics*, 132, 103978. <https://doi.org/10.1016/j.ijmedinf.2019.103978>

Abramski, K., Citraro, S., Lombardi, L., Rossetti, G., & Stella, M. (2023). Cognitive Network Science Reveals Bias in GPT-3, GPT-3.5 Turbo, and GPT-4 Mirroring Math Anxiety in High-School Students. *Big Data and Cognitive Computing*, 7(3). <https://doi.org/10.3390/bdcc7030124>

Agrawal, S. (2023). Are LLMs the Master of All Trades? : Exploring Domain-Agnostic Reasoning Skills of LLMs. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2303.12810>

Aher, G., Arriaga, R. I., & Kalai, A. T. (2023). *Using large language models to simulate multiple humans and*

*replicate human subject studies* Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA.

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2023). Playing repeated games with Large Language Models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.16867>

Ali, J. K. M., Shamsan, M. A. A., Hezam, T. A., & Mohammed, A. A. Q. (2023). Impact of ChatGPT on Learning Motivation. *Journal of English Studies in Arabia Felix*, 2(1), 41-49. <https://doi.org/10.56540/jesaf.v2i1.51>

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2022). Out of One, Many: Using Language Models to Simulate Human Samples. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2209.06899>

Atari, M., Xue, M. J., Park, P. S., Blasi, D. E., & Henrich, J. (2023). Which Humans? *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/5b26t>

Aydin, Ö., & Karaarslan, E. (2022). OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare. *Emerging Computer Technologies*(2), 22-31. <https://doi.org/10.2139/ssrn.4308687>

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada. <https://doi.org/10.1145/3442188.3445922>

Binz, M., & Schulz, E. (2023a). Turning large language models into cognitive models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2306.03917>

Binz, M., & Schulz, E. (2023b). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences of the United States of America*, 120(6), e2218523120. <https://doi.org/10.1073/pnas.2218523120>

Blyler, A. P., & Seligman, M. E. P. (2023a). AI assistance for coaches and therapists. *The Journal of Positive Psychology*, 1-13.

Blyler, A. P., & Seligman, M. E. P. (2023b). Personal narrative and stream of consciousness: an AI approach. *The Journal of Positive Psychology*, 1-7. <https://doi.org/10.1080/17439760.2023.2257666>

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., . . . Amodei, D. (2020). Language Models are Few-Shot Learners. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2005.14165>

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E. K., Kamar, E., Lee, P., Lee, Y. T., Li, Y.-F., Lundberg, S. M., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2303.12712>

Cai, Z. G., Haslett, D. A., Duan, X., Wang, S., & Pickering, M. J. (2023). Does ChatGPT resemble humans in language use? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2303.08014>

Carlbring, P., Hadjistavropoulos, H., Kleiboer, A., & Andersson, G. (2023). A new era in Internet interventions: The advent of Chat-GPT and AI-assisted therapist guidance. *Internet Interventions*, 32, 100621. <https://doi.org/10.1016/j.invent.2023.100621>

Chiang, C.-H., & Lee, H.-y. (2023). Can Large Language Models Be an Alternative to Human Evaluations? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.01937>

Crockett, M., & Messeri, L. (2023). Should large language models replace human participants? *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/4zdx9>

D'Oria, M. (2023). Can AI Language Models Improve Human Sciences Research? A Phenomenological Analysis and Future Directions. *Encyclopaideia*, 27(66), 77-92. <https://doi.org/10.6092/issn.1825-8670/16554>

De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21.  
<https://doi.org/10.1017/S1366728906002732>

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., Jones, M., Krettek-Cobb, D., Lai, L., Jones-Mitchell, N., Ong, D. C., Dweck, C. S., Gross, J. J., & Pennebaker, J. W. (2023). Using large language models in psychology. *Nature Reviews Psychology*, 2(11), 688-701. <https://doi.org/10.1038/s44159-023-00241-5>

Dergaa, I., Chamari, K., Zmijewski, P., & Ben Saad, H. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615-622. <https://doi.org/10.5114/biolsport.2023.125623>

Dhingra, S., Singh, M., Sb, V., Malviya, N., & Singh Gill, S. (2023). Mind meets machine: Unravelling GPT-4's cognitive psychology. *arXiv preprint*, arXiv:2303.11436. <https://doi.org/10.48550/arXiv.2303.11436>

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*, 27(7), 597-600. <https://doi.org/10.1016/j.tics.2023.04.008>

Elyoseph, Z., & Levkovich, I. (2023). Beyond human expertise- the promise and limitations of ChatGPT in suicide risk assessment. *Frontiers in Psychiatry*, 14. <https://doi.org/10.3389/fpsyg.2023.1213141>

Fecher, B., Hebing, M., Laufer, M., Pohle, J., & Sofsky, F. (2023). Friend or foe? Exploring the implications of large language models on the science system. *Ai & Society*. <https://doi.org/10.1007/s00146-023-01791-1>

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681-694. <https://doi.org/10.1007/s11023-020-09548-1>

Frank, M. C. (2023). Baby steps in evaluating the capacities of large language models. *Nature Reviews Psychology*, 2(8), 451-452. <https://doi.org/10.1038/s44159-023-00211-x>

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93.

Goertzel, B. (2023). Generative AI vs. AGI: The Cognitive Strengths and Weaknesses of Modern LLMs. *arXiv preprint*, arXiv:2309.10371. <https://doi.org/10.48550/arXiv.2309.10371>

Graber-Stiehl, I. (2023). IS THE WORLD READY FOR AI-POWERED THERAPY? *Nature*, 617, 22-24.  
<https://doi.org/10.1038/d41586-023-01473-4>

Grossmann, I., Feinberg, M., Parker, D. C., Christakis, N. A., Tetlock, P. E., & Cunningham, W. A. (2023). AI and the transformation of social science research. *Science*, 380(6650), 1108-1109.  
<https://doi.org/10.1126/science.ad1778>

Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. *IEEE Access*, 11, 80218-80245.  
<https://doi.org/10.1109/access.2023.3300381>

Hagendorff, T. (2023). Machine Psychology: Investigating Emergent Capabilities and Behavior in Large Language Models Using Psychological Methods. *arXiv preprint*.  
<https://doi.org/10.48550/arXiv.2303.13988>

Hagendorff, T., Fabi, S., & Kosinski, M. (2022). Machine intuition: Uncovering human-like intuitive decision-making in GPT-3.5. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2212.05206>

Hagendorff, T., Fabi, S., & Kosinski, M. (2023). Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science*, 3(10), 833-838.  
<https://doi.org/10.1038/s43588-023-00527-x>

Han, H. (2023). Potential Benefits of Employing Large Language Models in Research in Moral Education and Development. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2306.13805>

Harding, J., D'Alessandro, W., Laskowski, N. G., & Long, R. (2023). AI language models cannot replace human

research participants. *Ai & Society*. <https://doi.org/10.1007/s00146-023-01725-x>

Hardy, M., Sucholutsky, I., Thompson, B., & Griffiths, T. (2023). Large language models meet cognitive science: Llms as tools, models, and participants. Proceedings of the annual meeting of the cognitive science society,

Hayes, A. (2023). “Conversing” with Qualitative Data: Enhancing Qualitative Research through Large Language Models (LLMs). *PsyArXiv preprint*. <https://doi.org/10.31235/osf.io/yms8p>

Hendel, R., Geva, M., & Globerson, A. (2023). In-Context Learning Creates Task Vectors. *arXiv preprint*, arXiv:2310.15916. <https://doi.org/10.48550/arXiv.2310.15916>

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The Efficacy of Cognitive Behavioral Therapy: A Review of Meta-analyses. *Cognit Ther Res*, 36(5), 427-440. <https://doi.org/10.1007/s10608-012-9476-1>

Holtzman, A., West, P., & Zettlemoyer, L. (2023). Generative Models as a Complex Systems Science: How can we make sense of large language model behavior? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2308.00189>

Hothersall, D., & Lovett, B. J. (2022). *History of psychology*. Cambridge University Press.

Huang, J., & Chang, K. C.-C. (2022). Towards Reasoning in Large Language Models: A Survey. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2212.10403>

Hutson, M. (2023). Doing research with human subjects is costly and cumbersome. Can AI chatbots replace them? *Science*, 381(6654), 121-123. <https://doi.org/10.1126/science.adj6791>

Jin, C., Zhang, S., Shu, T., & Cui, Z. (2023). The Cultural Psychology of Large Language Models: Is ChatGPT a Holistic or Analytic Thinker? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2308.14242>

Jungherr, A. (2023). *Using ChatGPT and Other Large Language Model (LLM) Applications for Academic Paper Assignments*. <https://fis.uni-bamberg.de/handle/uniba/58950>  
urn:nbn:de:bvb:473-irb-589507

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103. <https://doi.org/10.1016/j.lindif.2023.102274>

Ke, L., Zhang, G., He, J., Li, Y., Li, Y., Liu, X., & Fang, P. (2023). Pilot Selection in the Era of Virtual Reality: Algorithms for Accurate and Interpretable Machine Learning Models. *Aerospace*, 10(5). <https://doi.org/10.3390/aerospace10050394>

Kjell, O., Kjell, K., & Schwartz, H. A. (2023). Beyond Rating Scales: With Care for Validation Large Language Models Are Poised to Change Psychological Assessment. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/yfd8g>

Kosinski, M. (2023). Theory of Mind May Have Spontaneously Emerged in Large Language Models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2302.02083>

Lamichhane, B. (2023). Evaluation of ChatGPT for NLP-based Mental Health Applications. *arXiv preprint*, arXiv:2303.15727. <https://doi.org/10.48550/arXiv.2303.15727>

Li, J., Tang, T., Zhao, W. X., Nie, J.-Y., & Wen, J.-R. (2022). Pretrained Language Models for Text Generation: A Survey. *arXiv preprint*, arXiv:2201.05273. <https://doi.org/10.48550/arXiv.2201.05273>

Li, M., Enkhtur, A., Cheng, F., & Yamamoto, B. A. (2023). Ethical implications of ChatGPT in higher education: A scoping review. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2311.14378>

Li, X., Li, Y., Liu, L., Bing, L., & Joty, S. (2022). Does GPT-3 Demonstrate Psychopathy? Evaluating Large

Language Models from a Psychological Perspective. *arXiv preprint*.  
<https://doi.org/10.48550/arXiv.2212.10529>

Liu, J. M., Li, D., Cao, H., Ren, T., Liao, Z., & Wu, J. (2023). ChatCounselor: A Large Language Models for Mental Health Support. *arXiv preprint*, arXiv:2309.15461. <https://doi.org/10.48550/arXiv.2309.15461>

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *AcM Computing Surveys*, 55(9), 1-35. <https://doi.org/10.1145/3560815>

Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., & Tang, J. (2022). P-Tuning: Prompt Tuning Can Be Comparable to Fine-tuning Across Scales and Tasks. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Dublin, Ireland.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of ChatGPT-Related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2).  
<https://doi.org/10.1016/j.metrad.2023.100017>

Loconte, R., Orrù, G., Tribastone, M., Pietrini, P., & Sartori, G. (2023). Challenging ChatGPT's "intelligence" with human tools: A Neuropsychological Investigation on Prefrontal Functioning of a Large Language Model. *SSRN preprint*. <https://doi.org/10.2139/ssrn.4471829>

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). Dissociating language and thought in large language models: a cognitive perspective. *arXiv preprint*.  
<https://doi.org/10.48550/arXiv.2301.06627>

Marjieh, R., Sucholutsky, I., Rijn, P. v., Jacoby, N., & Griffiths, T. L. (2023). Large language models predict human sensory judgments across six modalities. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2302.01308>

Miotto, M., Rossberg, N., & Kleinberg, B. (2022). Who is GPT-3? An Exploration of Personality, Values and Demographics. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2209.14338>

Mitchell, M. (2023). AI's challenge of understanding the world. *Science*, 382(6671).  
<https://doi.org/10.1126/science.adm8175>

Nazario-Johnson, L., Zaki, H. A., & Tung, G. A. (2023). Use of large language models to predict neuroimaging. *Journal of the American College of Radiology*, 20(10), 1004-1009.  
<https://doi.org/10.1016/j.jacr.2023.06.008>

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: holistic versus analytic cognition. *Psychological Review*, 108(2), 291-310. <https://doi.org/10.1037/0033-295X.108.2.291>

OpenAI. (2023). GPT-4 Technical Report. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2303.08774>

Orrù, G., Piarulli, A., Conversano, C., & Gemignani, A. (2023). Human-like problem-solving abilities in large language models using ChatGPT. *Frontiers in Artificial Intelligence*, 6, 1199350.  
<https://doi.org/10.3389/frai.2023.1199350>

Pal, R., Garg, H., Patel, S., & Sethi, T. (2023). Bias Amplification in Intersectional Subpopulations for Clinical Phenotyping by Large Language Models. *medRxiv preprint*. <https://doi.org/10.1101/2023.03.22.23287585>

Park, B., & Judd, C. M. (2005). Rethinking the Link Between Categorization and Prejudice Within the Social Cognition Perspective. *Personality and Social Psychology Review*, 9(2), 108-130.  
[https://doi.org/10.1207/s15327957pspr0902\\_2](https://doi.org/10.1207/s15327957pspr0902_2)

Park, J. S., Popowski, L., Cai, C., Morris, M. R., Liang, P., & Bernstein, M. S. (2022). Social simulacra: Creating populated prototypes for social computing systems. Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology,

Park, P. S., Schoenagger, P., & Zhu, C. (2023). Diminished Diversity-of-Thought in a Standard Large Language Model. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2302.07267>

Patel, S. C., & Fan, J. (2023). Identification and Description of Emotions by Current Large Language Models. *bioRxiv preprint*. <https://doi.org/10.1101/2023.07.17.549421>

Peng, Y., Han, J., Zhang, Z., Fan, L., Liu, T., Qi, S., Feng, X., Ma, Y., Wang, Y., & Zhu, S.-C. (2023). The Tong Test: Evaluating Artificial General Intelligence Through Dynamic Embodied Physical and Social Interactions. *Engineering*. <https://doi.org/10.1016/j.eng.2023.07.006>

Peters, H., & Matz, S. (2023). Large Language Models Can Infer Psychological Dispositions of Social Media Users. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2309.08631>

Qureshi, R., Shaughnessy, D., Gill, K. A. R., Robinson, K. A., Li, T., & Agai, E. (2023). Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? *Systematic Reviews*, 12(1), 72. <https://doi.org/10.1186/s13643-023-02243-z>

Rathje, S., Mirea, D.-M., Sucholutsky, I., Marjeh, R., Robertson, C., & Bavel, J. J. V. (2023). GPT is an effective tool for multilingual psychological text analysis. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/sekf5>

Salah, M., Al Halbusi, H., & Abdelfattah, F. (2023). May the force of text data analysis be with you: Unleashing the power of generative AI for social psychology research. *Computers in Human Behavior: Artificial Humans*, 1(2). <https://doi.org/10.1016/j.chbah.2023.100006>

Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)*, 11(6). <https://doi.org/10.3390/healthcare11060887>

Sap, M., LeBras, R., Fried, D., & Choi, Y. (2022). Neural Theory-of-Mind? On the Limits of Social Intelligence in Large LMs. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2210.13312>

Sartori, G., & Orrù, G. (2023). Language models and psychological sciences. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1279317>

Schaaff, K., Reinig, C., & Schlippe, T. (2023). Exploring ChatGPT's Empathic Abilities. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2308.03527>

Schueller, S. M., & Morris, R. R. (2023). Clinical science and practice in the age of large language models and generative artificial intelligence. *Journal of Consulting and Clinical Psychology*, 91(10), 559-561. <https://doi.org/10.1037/ccp0000848>

Seals, S. M., & Shalin, V. L. (2023). Long-form analogies generated by chatGPT lack human-like psycholinguistic properties. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2306.04537>

Sejnowski, T. (2022). Large Language Models and the Reverse Turing Test. *arXiv preprint*. <https://doi.org/10.48550/arxiv.2207.14382>

Sha, H., Mu, Y., Jiang, Y., Chen, L., Xu, C., Luo, P., Eben Li, S., Tomizuka, M., Zhan, W., & Ding, M. (2023). LanguageMPC: Large Language Models as Decision Makers for Autonomous Driving. *arXiv preprint*, arXiv:2310.03026. <https://doi.org/10.48550/arXiv.2310.03026>

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1), 46-57. <https://doi.org/10.1038/s42256-022-00593-2>

Shiffrin, R., & Mitchell, M. (2023). Probing the psychology of AI models. *Proceedings of the National Academy of Sciences of the United States of America*, 120(10), e2300963120. <https://doi.org/10.1073/pnas.2300963120>

Simon, H. A. (1979). Information Processing Models of Cognition. *Annual Review of Psychology*, 30(1), 363-

396. <https://doi.org/10.1146/annurev.ps.30.020179.002051>

Stade, E. C., Stirman, S. W., Ungar, L., Boland, C. L., Schwartz, H. A., Yaden, D. B., Sedoc, J., Derubeis, R. J., Willer, R., & Eichstaedt, J. C. (2023). Large Language Models Could Change the Future of Behavioral Healthcare: A Proposal for Responsible Development and Evaluation. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/cuzvr>

Stella, M., Hills, T. T., & Kenett, Y. N. (2023). Using cognitive psychology to understand GPT-like models needs to extend beyond human biases. *Proceedings of the National Academy of Sciences of the United States of America*, 120(43), e2312911120. <https://doi.org/10.1073/pnas.2312911120>

Stevenson, C., Smal, I., Baas, M., Grasman, R., & Maas, H. v. d. (2022). Putting GPT-3's Creativity to the (Alternative Uses) Test. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2206.08932>

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: an autoethnographic study. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00404-7>

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays — should professors worry? *Nature*. <https://doi.org/10.1038/d41586-022-04397-7>

Suri, G., Slater, L. R., Ziaeef, A., & Nguyen, M. (2023). Do Large Language Models Show Decision Heuristics Similar to Humans? A Case Study Using GPT-3.5. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.04400>

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, 33(1), 1-39.

Talboy, A. N., & Fuller, E. (2023). Challenging the appearance of machine intelligence: Cognitive bias in LLMs. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2304.01358>

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2102.02503>

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine*, 29(8), 1930-1940. <https://doi.org/10.1038/s41591-023-02448-8>

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models know what humans know? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2209.01515>

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature*, 614(7947), 224-226. <https://doi.org/10.1038/d41586-023-00288-7>

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P., Deac, A., Anandkumar, A., Bergen, K., Gomes, C. P., Ho, S., Kohli, P., Lasenby, J., Leskovec, J., Liu, T. Y., Manrai, A., . . . Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972), 47-60. <https://doi.org/10.1038/s41586-023-06221-2>

Wang, X., Li, X., Yin, Z., Wu, Y., & Liu, J. (2023). Emotional intelligence of Large Language Models. *Journal of Pacific Rim Psychology*, 17. <https://doi.org/10.1177/18344909231213958>

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, 7(9), 1526-1541. <https://doi.org/10.1038/s41562-023-01659-w>

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent Abilities of Large Language Models. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2206.07682>

Yildirim, I., & Paul, L. A. (2023). From task structures to world models: What do LLMs know? *arXiv preprint*,

arXiv:2310.04276. <https://doi.org/10.48550/arXiv.2310.04276>

Yukun, Z., Xu, L., Huang, Z., Peng, K., Seligman, M., Li, E., & Yu, F. (2023). AI chatbot responds to emotional cuing. *PsyArXiv preprint*. <https://doi.org/10.31234/osf.io/9ymfz>

Zhang, J., Xu, X., & Deng, S. (2023). Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View. *arXiv preprint*, arXiv:2310.02124. <https://doi.org/10.48550/arXiv.2310.02124>

Zhong, Y., Chen, Y. J., Zhou, Y., Lyu, Y. A., Yin, J. J., & Gao, Y. J. (2023). The Artificial intelligence large language models and neuropsychiatry practice and research ethic. *Asian Journal of Psychiatry*, 84, 103577. <https://doi.org/10.1016/j.ajp.2023.103577>

Zhuang, Y., Liu, Q., Ning, Y., Huang, W., Lv, R., Huang, Z., Zhao, G., Zhang, Z., Mao, Q., Wang, S., & Chen, E. (2023). Efficiently Measuring the Cognitive Ability of LLMs: An Adaptive Testing Perspective. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2306.10512>

Ziems, C., Held, W., Shaikh, O., Chen, J., Zhang, Z., & Yang, D. (2023). Can Large Language Models Transform Computational Social Science? *arXiv preprint*. <https://doi.org/10.48550/arXiv.2305.03514>